

# CHAPTER 17

## SCORE REPORTING

Table 17-1 lists the primary MEA reports.

Table 17-1 Primary MEA Reports	
1.	<i>Student Report for Parents/Guardians</i>
2.	<i>Student Labels</i>
3.	<i>School Common Item Level Class Report</i>
4.	<i>School Report</i>
5.	<i>District Report</i>
6.	<i>Student Writing CD</i>

### **STUDENT REPORT FOR PARENTS/GUARDIANS**

Student reports show the scaled score for each subject area, as well as a score band that indicates the standard error of measurement surrounding each score. Performance level definitions are provided so that parents/guardians will understand how to interpret the scaled scores. Specific comments are provided about the student's writing performance. Information is also provided to show how the student's performance compared to the average scores from the student's school, district, and state. An overview of test content is provided, along with a cautionary statement about interpreting scores and guidelines for parents/guardians for helping their children improve.

### **STUDENT LABELS**

To aid schools in keeping track of student scores, schools were supplied with student score information on individual labels that they could affix to files, if desired.

### **SCHOOL COMMON ITEM LEVEL CLASS REPORT**

The *Common Item Level Class Report* shows the answers that each student gave on the multiple-choice questions, as well as his/her score on each open-response question. The report also summarizes overall performance at the school, district, and state levels for each of the question types.

## **SCHOOL AND DISTRICT REPORTS**

The school and district reports are intended for administrators and other interested parties. The school report includes performance level definitions, scaled score intervals, and information about how summary statistics are affected by students not tested; all of which are intended to help the reader interpret the report. The school report provides all results for the school, the district, and the entire state. The results provided are

- the number of students tested by student status (regular, students with disabilities, and limited English proficient students) for all subject areas combined and separately for each subject area,
- the percentage of students in each performance level by subject area,
- the distribution of scaled scores by subject area,
- the number of students in each performance level by subject area and student status,
- subject area subscores outlining the number of possible points by learning results standards,
- three-year comparisons of school results, and
- average subject score by number of years in the school or district.

The district report is the same as the school report, except that it does not include the school-level data and the three-year comparisons are by district rather than by school.

## **STUDENT WRITING CD**

The student writing CD contains all of the student's writing for each school. The schools are then able to print out and/or review the actual student's work.

Sample reports can be found in Appendix C.

# QUALITY CONTROL (QC) PROCESS FOR ENSURING ACCURACY OF PRINTED REPORTS

## **GENERAL**

1. Whenever new reports are received from Measurement, Design, and Analysis (MDA), the *date and time* they were received is written at the top of each report, so that it will be easy to identify the most recent version of each report.
2. For each of the items that follow, a checkmark was put in a logical position on each report to indicate that each check was done. For instance, after verifying that a name is correct, a checkmark is placed next to the name; after verifying that a score is correct or that a bar is the correct length, a checkmark is placed next to it; and so on. This lets other QC staff verify which checks have been done and which have not.
3. When all checks are completed on a given report, the QC staff's initials and that day's date are written at the top of the page so that everyone knows who checked them.

## **PARENT REPORTS:**

### **Letter Side:**

1. Proofing text and formatting of entire side is done once thoroughly, and then spot-checked in additional QC runs.
2. The State MEA Summary Results (bottom right box): the percentages are verified that they match those on the school and district reports for the state (page 2 bar graph, page 4, and page 6). The bars are then checked to make sure that they accurately represent the percentages reported.

### **Performance Assessment Side:**

1. Proofing text and formatting of entire side is done once thoroughly, and then spot-checked in additional QC runs.
2. It is verified that the student name and grade are the same as those printed on the letter side.
3. QC staff also checks to make sure that the performance level corresponds to the scaled score.

4. They also make sure that the diamond placement in the top box corresponds to score and performance levels and that the range bar does not fall outside of the scale area.
5. If the student was excluded or testing was incomplete, it was verified that no scaled score or performance level appeared, nor were there diamonds or range bars. Instead, it was verified that the words “*excluded*” or “*testing incomplete*” appeared in the performance level box.
6. The performance level and scaled score was compared to the common item report to ensure that they matched. They were also compared to the labels to ensure that they matched.
7. It was verified that the school, district, and state averages matched those in the school and district reports (page 2). And they also verified the accuracy of the height of the bars. To make sure that the height of the bar reflected the number on top of the bar, QC staff looked to the left of the bars at the scale. (The bar height should match the performance level.)
8. It was verified that there were no student bars, if a student was excluded or testing was incomplete. (Instead, he/she would get the school, district, and state bars only.)
9. Writing comments were checked that the commendations/needs correspond to the comment codes on the Common Item Class Report (for individual students). It was also checked that the comments were properly categorized (e.g., needs statements into Needs box and commendations statements into Commendation box).
10. Students marked as **NT** (not tested) or **TI** (tested incomplete) *may* still have comments. It was verified that any comments matched what was reported on the Common Item Level Report.
11. Student’s Performance in Content Area subcategories: Diamond placement was verified. It was checked that the diamonds did *not* overlap borders, nor did their corners get cut off. It was also checked that there was a diamond for each of the three categories. And if a student was *excluded* or the testing was *incomplete*, then there were *no* diamonds. (If a student had scores for Writing but was incomplete or excluded for Reading, then the diagram would show two diamonds in the Writing category, but no diamond for Reading.)

## ***LABELS***

1. Spelling, punctuation, and formatting (for margins, fit of text on the label, and so on) were checked.
2. It was verified that the school and district information is correct.
3. The names, proficiency levels, and scaled scores were checked to make sure they matched what was reported on the common item and parent reports.
4. It was also verified that the students listed as belonging to a given school were the same on all reports for that school.
5. The grade was verified and that each page of labels includes information for only *one* school.

## ***COMMON ITEM CLASS REPORTS (READING AND WRITING)***

1. The QC staff was directed to proof the text and formatting of report, including legend (on reverse side), if provided.
2. They also compared the heading information to the shells and verified that the data in the heading matched the data in the shells.
3. They then verified that the names appear in alphabetical order, and in groups of five.
4. The staff was then told to **Highlight** the information for any student who was *excluded or incomplete* (marked with **asterisks**). Subtract these students from the total and indicate the new total next to the original “group size” indicated in the box at the top of the page (this is the number used when calculating averages).
5. It was verified that the number of points per score did not exceed the maximum value indicated in the heading. (If the number 4 is written in the total possible points box then no one should have an 8 for a score.)
6. The keys were then verified by comparing each correct answer to the incorrect answers listed underneath for each question. (For example, if A is the correct MC, there should be no A’s for incorrect answers.)
7. Next, the number of students receiving each *type* of annotation was counted. A need or commendation with the same 1<sup>st</sup> letter should only be counted once per student. (For example, a student who received two *needs*

that began with a T [for “Topic Development”] and one *commendation* that began with a T, would only be counted *once* for the *needs* and *once* for the *commendation*.) These numbers should match those reported on page 8 of the school and district reports.

8. Then the QC staff calculates the average scaled score and the average points earned for the school.
9. Finally, the match to school and district reports takes place by adding across classes to get school scores, and across schools to get district scores (remember to skip the highlighted students and divide by the adjusted group size).

$$\frac{\text{Total of all scaled scores}}{\text{Total number of students}} = \text{average scaled score for the class}$$

$$\frac{\text{Total of all points earned}}{\text{Total number of students}} = \text{average points earned for the class}$$

### ***SCHOOL AND DISTRICT REPORTS***

#### **Page 1:**

1. The entire page was proofed for both text and formatting errors, including verifying the page references in the table of contents.

#### **Page 2:**

1. The entire page was proofed for both text and formatting errors once thoroughly.
2. It was verified that the scaled scores matched the ones on the parent report and the state-score handout (provided by MDA).
3. The percentage tables were then checked to make sure that the state percentages matched those on the parent reports and handout. The school and district should match the percentages on page 4, 6, or 9.
4. The scores reported for the school and district under Average Performance Score to the averages calculated from the common item reports were compared next.
5. Then the staff calculated and verified the accuracy of the Cum. Avg. under Average Performance Score. Total both averages for last year and this year and divide by two.
6. Finally, they compared this last year’s reports to verify historic data.

**Page 3:**

1. The entire page was proofed for both text and formatting errors. The informational paragraph at the top of the page was checked so that it refers to school or district as appropriate.
2. It was verified that the students enrolled on (school report and district report) equaled the number(s) listed as group size on the common item report.
3. All percentages (except the last two rows) were computed by taking the number in each row and dividing it by the number enrolled.

**Pages 4, 6, and 9 (Reading and Writing) and pages 4, 6, 8, and 11 (Mathematics, Science & Tech., and Social Studies, and VPA):**

1. All pages were proofread for both text and formatting errors once thoroughly.
2. QC staff added up the number of students at each performance level (school “N” and district “N”) to get the total included for that content area. And it was verified that it matched the number of students on the common item report (the modified total, minus excluded and incomplete students).
3. Then the percent of students at each level was verified by dividing the number at that level by the total number of students included for that content area. Add the percents down the levels to make sure they equal 99-101.
4. Under “Average Points Attained,” the percentage for school, district, and state was verified by dividing the number (“N”) by the number of points possible. (Note: If the school or district is small, *some of these cells may be blank*. Each Learning Results Content Standard must have *at least 5 students* to be reported in this table.)

**Pages 5, 7, and 10 (Reading and Writing) and pages 5, 7, 9, and 11 (Mathematics, Science & Tech., Social Studies, and VPA):**

1. All pages were proofread for both text and formatting errors once thoroughly.
2. It was verified that the percentages for each option equaled 99-101 per question.
3. QC staff then checked percentages for reasonableness. (If the total *number of students* in a category is *less than 5*, no percentage will be reported. Percentages for “special” categories, such as “Migrant,” might total less than 100.)

**Page 8: Summary of Annotations Table (Writing Only):**

1. The entire page was proofread for both text and formatting errors once thoroughly.
2. It was verified that the number of students receiving a commendation or need matched the number counted on the common item report. A need or commendation with the same 1<sup>st</sup> letter was only counted once per student. [For example, a student who received two *needs* that began with a T (for “Topic Development”) and one *commendation* that began with a T, would only be counted *once* for the *needs* and *once* for the *commendation*.]
3. Staff then recalculated the percentages by dividing the number of students reported in this table by the total number of students tested in writing for the school and/or district. To get the number tested in writing, add up the number of students at each performance level on page six.

## **REPORTING IRREGULARITIES**

In the subgroup results section of the school and district reports, there were lines for students with identified disabilities and all other students. The wrong group of students with disabilities was used for this analysis. For that reason, corrected data were sent to affected schools.



# REFERENCES

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1985). *Standards for educational and psychological testing*. Washington, DC: APA.
- Brown, F. G. (1983). *Principles of educational and psychological testing* (3rd Edition). Fort Worth: Holt, Rinehart and Winston.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.
- Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.
- Dorans, N. J., & Holland, P.W. (1993). DIF detection and description. In P.W. Holland & H. Wainer (Eds.) *Differential item functioning* (pp. 35–66). Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Dorans, N.J., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement*, 23, 355–368.
- Kingston, N.M., Kahl, S.R., Sweeney, K.P., & Bay, L. (2001). Setting performance standards using the Body of Work (BoW) method. In G.J. Cizek (Ed.) *Setting Performance Standards: Concepts, Methods, and Perspectives* (pp. 219-248). Mahwah, NJ: Lawrence Earlbaum and Associates.
- Livingston, S.A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, 32, 179–197.
- Livingston, S.A., & Zieky, M.J. (1982). *Passing scores*. Princeton, NJ: Educational Testing Service.
- Medsker, G. J. (1998). *Determining the relationship between student transience and KIRIS school results: are schools with transient students unfairly impacted?* (HumRRO Report, FR-WATSD-98-12). Alexandria, VA: Human Resources Research Organization.
- Feldt, L.S., & Brennan, R.L. (1989). Reliability. In R.L. Linn (Ed.) *Educational Measurement* (3<sup>rd</sup> Ed. Pp.105-146).